

Ensemble Data Assimilation Without Perturbed Observations

Jeffrey S. Whitaker* and Thomas M. Hamill

NOAA-CIRES Climate Diagnostics Center, Boulder, CO

submitted to *Mon. Wea. Rev.*

17th December 2001

*NOAA Climate Diagnostics Center, 325 Broadway R/CDC1, Boulder, CO 80305-3328, jsw@cdc.noaa.gov

Abstract

The ensemble Kalman filter (EnKF) is a data assimilation scheme based on the traditional Kalman-filter update equation. An ensemble of forecasts are used to estimate the background-error covariances needed to compute the Kalman gain. It is known that if the same observations and the same gain are used to update each member of the ensemble, the ensemble will systematically underestimate analysis-error covariances. This will cause a degradation of subsequent analyses and may lead to filter divergence. For large ensembles, it is known that this problem can be alleviated by treating the observations as random variables, adding random perturbations to them with the correct statistics.

Two important consequences of sampling error in the estimate of analysis-error covariances in the EnKF are discussed here. The first results from the analysis-error covariance being a nonlinear function of the background-error covariance in the Kalman filter. Due to this nonlinearity, analysis-error covariance estimates may be negatively biased, even if the ensemble background-error covariance estimates are unbiased. This problem must be dealt with in any Kalman-filter based ensemble data assimilation scheme.

A second consequence of sampling error is particular to schemes like the EnKF that use perturbed observations. While this procedure gives asymptotically correct analysis-error covariance estimates for large ensembles, the addition of perturbed observations adds an additional source of sampling error related to the estimation of the observation-error covariances. In addition to reducing the accuracy of the analysis-error covariance estimate, this extra source of sampling error increases the probability that the analysis-error covariance will be underestimated. Because of this, ensemble data assimilation

methods which use perturbed observations are expected to be less accurate than those which do not.

Several ensemble filter formulations have recently been proposed which do not require perturbed observations. Here we examine a particularly simple implementation which we call the “ensemble square-root filter,” or EnSRF. The EnSRF uses the traditional Kalman gain for updating the ensemble mean but uses a “reduced” Kalman gain to update deviations from the ensemble mean. There is no additional computational cost incurred by the EnSRF relative to the EnKF when the observations have independent errors and are processed one at a time. Using a hierarchy of perfect model assimilation experiments, we demonstrate that the elimination of the sampling error associated with the perturbed observations makes the EnSRF more accurate than the EnKF for the same ensemble size.

1. Introduction

The ensemble Kalman filter (EnKF), introduced by Evensen (1994), is a Monte-Carlo approximation to the traditional Kalman filter (KF) (Kalman and Bucy 1961; Gelb et al. 1974). The EnKF uses an ensemble of forecasts to estimate background-error covariances. By providing flow- and location-dependent estimates of background error, the EnKF can more optimally adjust the background to newly available observations. In doing so, it may be able to produce analyses and forecasts that are much more accurate than current operational data assimilation schemes which assume that the background error is known a priori and does not vary in time. The EnKF automatically provides a random sample of the estimated analysis-error distribution, which can then be used as initial conditions for an ensemble prediction system. The EnKF and four-dimensional variational data assimilation (4D-Var) are presently the two leading candidates to replace the current generation of three-dimensional variational (3D-Var) schemes (see Hamill et al. (2001) and references therein for a discussion of the merits of each approach).

Houtekamer and Mitchell (1998) recognized that in order for the EnKF to maintain sufficient spread in the ensemble and prevent filter divergence, the observations should be treated as random variables. They introduced the concept of using perturbed sets of observations to update each ensemble member. The perturbed observations consisted of the actual, or “control” observations plus random noise, with the noise randomly sampled from the observational-error distribution used in the data assimilation. Different ensemble members were updated using different sets of perturbed observations. Burgers et al. (1998) provided a theoretical justification for perturbing the observations in the EnKF, showing that if the ob-

servations are not treated as random variables, analysis error covariances are systematically underestimated, typically leading to filter divergence. Interestingly, in early implementations of the EnKF, perturbed observations (e.g. Evensen 1994; Evensen and van Leeuwen 1996) were not used, yet problems with filter divergence were not found. Most likely this was because for these oceanographic applications, observation errors were typically much smaller than either background errors or the system noise. The relative influence of noise associated with observational errors is much greater in atmospheric applications.

In this paper we discuss two important consequences by sampling error in ensemble data assimilation schemes based on the Kalman filter. These are associated with (1) the nonlinear dependence of the analysis-error covariance estimate on the background-error covariance estimate, and (2) the noise added to the observations. The former is a property of all Kalman filter based schemes, while the latter is specific to algorithms like the EnKF requiring perturbed observations. The consequences of the nonlinear dependence of the analysis-error covariance estimate on the background-error covariance estimate are similar to the “inbreeding” effect noted by Houtekamer and Mitchell (1998) and analyzed by van Leeuwen (1999) and Houtekamer and Mitchell (1999). The consequences of perturbing of observations have not been thoroughly explored in previous critiques of the EnKF (e.g. Anderson 2001; van Leeuwen 1999).

Ensemble filtering approaches have been designed without perturbed observations (e.g. Lermusiaux and Robinson 1999; Anderson 2001; Bishop et al. 2001). Here, we propose another ensemble filtering algorithm that does not require the observations to be perturbed, is conceptually simple, and, if observations are processed one at a time, is just as fast as

the EnKF. We will demonstrate that this new algorithm, which we shall call the "ensemble square-root filter" (EnSRF) is more accurate than the EnKF for a given ensemble size.

The remainder of the paper will be organized as follows: section 2 provides a background on the formulation of the EnKF as well as a simple example of how nonlinearity in the covariance update and noise from perturbed observations can cause problems in ensemble data assimilation. Section 3 presents the EnSRF algorithm, and sections 4 and 5 discuss comparative tests of the EnKF and EnSRF in a low-order model and a simplified global circulation model (GCM) respectively.

2. Background

a. Formulation of the ensemble Kalman filter

Following the notation of Ide et al. (1997), let \mathbf{x}^b be an m -dimensional background model forecast; let \mathbf{y}^o be an p -dimensional set of observations; let \mathbf{H} be the operator that converts the model state to the observation space (here, assumed linear, though it need not be); let \mathbf{P}^b be the $m \times m$ -dimensional background error covariance matrix; and let \mathbf{R} be the $p \times p$ -dimensional observation-error covariance matrix. The minimum error-variance estimate of the analyzed state \mathbf{x}^a is then given by the traditional Kalman filter update equation (Lorenc 1986),

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b), \quad (1)$$

where

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T (\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R})^{-1}. \quad (2)$$

The analysis error covariance \mathbf{P}^a is reduced by the introduction of observations by an amount given by

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^b(\mathbf{I} - \mathbf{KH})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T = (\mathbf{I} - \mathbf{KH})\mathbf{P}^b. \quad (3)$$

The derivation of Eq. (3) assumes $\mathbf{P}^a = \langle (\mathbf{x}^t - \mathbf{x}^a)(\mathbf{x}^t - \mathbf{x}^a)^T \rangle$, where \mathbf{x}^t is the true state and $\langle \cdot \rangle$ denotes the expected value. The definition of \mathbf{x}^a is then substituted from (1), and observation and background errors are assumed to be uncorrelated.

In the EnKF, \mathbf{P}^b is approximated using the sample covariance from an ensemble of model forecasts. For the rest of the paper, the symbol \mathbf{P} is used to denote the sample covariance from an ensemble, and \mathbf{K} is understood to be computed using sample covariances. Expressing the variables as an ensemble mean (denoted by an over-bar) and a deviation from the mean (denoted by a prime), the update equations for the EnKF may be written as

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{K}(\bar{\mathbf{y}}^o - \mathbf{H}\bar{\mathbf{x}}^b), \quad (4)$$

$$\mathbf{x}'^a = \mathbf{x}'^b + \tilde{\mathbf{K}}(\mathbf{y}'^o - \mathbf{H}\mathbf{x}'^b), \quad (5)$$

where $\mathbf{P}^b = \overline{\mathbf{x}'^b \mathbf{x}'^{bT}} \equiv \frac{1}{n-1} \sum_{i=1}^n \mathbf{x}'^b_i \mathbf{x}'^{bT}_i$, n is the ensemble size, \mathbf{K} is the traditional Kalman gain given by Eq. (2) and $\tilde{\mathbf{K}}$ is the gain used to update deviations from the ensemble mean. In the EnKF, $\tilde{\mathbf{K}} = \mathbf{K}$, and \mathbf{y}'^o are randomly drawn from the probability distribution of observation errors (Burgers et al. 1998). Note that wherever an overbar is used in the context of a covariance estimate a factor of $n - 1$ instead of n is implied in the denominator, so that the estimate is unbiased. In the EnKF framework, there is no need to compute and store the full

matrix \mathbf{P}^b . Instead, $\mathbf{P}^b \mathbf{H}^T$ and $\mathbf{H} \mathbf{P}^b \mathbf{H}^T$ are estimated directly using the ensemble (Evensen 1994; Houtekamer and Mitchell 1998).

In the EnKF, if all members are updated with the same observations ($\mathbf{y}'^o = \mathbf{0}$) using the same gain ($\mathbf{K} = \tilde{\mathbf{K}}$), the covariance of the analyzed ensemble can be shown to be

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b(\mathbf{I} - \mathbf{K}\mathbf{H})^T \quad (6)$$

(Burgers et al. 1998). The missing term $\mathbf{K}\mathbf{R}\mathbf{K}^T$ causes \mathbf{P}^a to be systematically underestimated. If random noise is added to the observations so that $\mathbf{y}'^o \neq \mathbf{0}$, the analyzed ensemble variance is

$$\begin{aligned} \mathbf{P}^a = & (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \\ & \mathbf{K}(\overline{\mathbf{y}'^o \mathbf{y}'^o{}^T} - \overline{\mathbf{H}\mathbf{x}'^b \mathbf{y}'^o{}^T} - \overline{\mathbf{y}'^o \mathbf{x}'^b{}^T \mathbf{H}^T})\mathbf{K}^T + \\ & \overline{\mathbf{x}'^b \mathbf{y}'^o{}^T} \mathbf{K}^T + \mathbf{K} \overline{\mathbf{y}'^o \mathbf{x}'^b{}^T}. \end{aligned} \quad (7)$$

If the observation noise is defined such that $\overline{\mathbf{y}'^o \mathbf{y}'^o{}^T} = \mathbf{R}$ the expected value of \mathbf{P}^a is equal to that traditional Kalman filter result (Eq. 3), since the expected value of the background-observation error covariance $\overline{\mathbf{x}'^b \mathbf{y}'^o{}^T}$ is zero (Burgers et al. 1998).

b. The impact of sampling error on analysis error covariance estimates

For a finite-sized ensemble, there is sampling error in the estimation of background-error covariances. In other words, $\mathbf{P}_\infty \neq \mathbf{P}_n$, where \mathbf{P}_n is the error covariance obtained from an n member ensemble and \mathbf{P}_∞ is the “exact” error covariance, defined to be that which would be obtained from an infinite ensemble. When observations are perturbed in the EnKF,

there is also sampling error in the estimation of the observation-error covariances. We now explore how these can result in a mis-estimation of the analysis-error covariance in a simple experiment with a small ensemble.

To wit, consider an ensemble data assimilation system updating a 1-dimensional state vector \mathbf{x}^b to a new observation \mathbf{y}^o . One million multiple replications of a simple experiment are performed, an experiment where we are interested solely in the accuracy of the analysis-error covariance. \mathbf{x}^b and \mathbf{y}^o are assumed to represent the same quantity, so $\mathbf{H} = \mathbf{1}$. In each replication of the experiment, a five-member ensemble of \mathbf{x}^b is created, with each ensemble member sampled randomly from a $N(0,1)$ distribution. The distribution of background-error variance estimates computed for these ensembles form a chi-square distribution with four degrees of freedom (denoted by the solid curve in Fig. 1a). Since these estimates are unbiased, the mean of this distribution is equal to the true value ($\mathbf{P}^b = 1$). Two types of data assimilation methods are then tested. The first is the EnKF, in which each member is updated to an associated perturbed observation. Assuming $\mathbf{R} = \mathbf{1}$, the perturbed observations are randomly sampled from a $N(0,1)$ distribution and then adjusted so each set of perturbed observations has zero mean and unit variance. The second is a hypothetical ensemble filter that does not require perturbed observations (EnKF-noPO). For the EnKF-noPO, once the background-error covariance is estimated from the randomly sampled ensemble, we simply assume that the analysis-error covariance can be predicted by Eq. 3.

Let us first illustrate the consequences of sampling error associated with the nonlinear dependence of the analysis-error covariance on the background-error covariance. After the first observation, Eq. 3 simplifies to $\mathbf{P}^a = (1 - \mathbf{K})\mathbf{P}^b = (1 - \gamma\mathbf{P}^b)\mathbf{P}^b$, where $\gamma = 1/(\mathbf{P}^b + \mathbf{R})$.

Because of sampling error, \mathbf{P}^b is a random variable, so \mathbf{P}^a is a random variable as well, even in the absence of perturbed observations. The distribution of \mathbf{P}^a after the assimilation of a single observation in the EnKF-noPO is shown by the dashed curve in Fig. 1a. As expected, the effect of the observation is to reduce the ensemble variance, as evidenced by the higher probabilities at lower values of the error variance for \mathbf{P}^a (dashed curve), as compared to \mathbf{P}^b (solid curve). For this simple scalar example, Eq. 3 is akin to a power transformation applied to the sampling distribution for \mathbf{P}^b (e.g. Wilks 1995, section 3.4). This particular transformation reduces \mathbf{P}^a associated with large \mathbf{P}^b much more than it reduces it for small \mathbf{P}^b . In this case, the expected value for \mathbf{P}^b is 1.0, and the expected value of \mathbf{P}^a is 0.5. However, the mean of the \mathbf{P}^a distribution is ~ 0.44 , indicating the distribution of \mathbf{P}^a is biased even though the distribution of \mathbf{P}^b was not. This effect is related to the so-called “inbreeding” problem noted by Houtekamer and Mitchell (1998) and analyzed by van Leeuwen (1999).

Since this effect is purely a result of sampling error in the estimation of \mathbf{P}^b , and not perturbing the observations, all ensemble data assimilation schemes based on the Kalman filter should suffer from its consequences, not just the EnKF. For a finite ensemble, ensemble-mean errors will, on average, be larger than the “exact” value of \mathbf{P}^a . Since the bias associated with the nonlinear dependency of \mathbf{K} on \mathbf{P}^b results in a systematic underestimation of “exact” \mathbf{P}^a , the ensemble analysis-error covariance will systematically underestimate the actual ensemble mean analysis error. In ensemble-based data assimilation systems, if the analysis-error covariances systematically underestimate ensemble mean analysis error, observations will be under-weighted, and filter divergence may occur. This is why methods for increasing ensemble variance, such as the covariance inflation technique described by Anderson and Anderson

(1999) and the “double EnKF” proposed by Houtekamer and Mitchell (1998), are necessary.

Problems caused by the non-linearity in the covariance update (3) should lessen for larger ensembles. In that case, the sampling distribution for \mathbf{P}^b becomes narrower and more symmetric about its mean value, and the power transformation associated with the nonlinearity in the Kalman gain results in a \mathbf{P}^a distribution which is itself more symmetric and less biased.

We now consider the effect of perturbing the observations on analysis-error covariance estimates. For the EnKF, \mathbf{P}^a is given by Eq. 7, which in this experiment simplifies to

$$\begin{aligned}\mathbf{P}^a &= (1 - \mathbf{K})^2 \mathbf{P}^b + \mathbf{K}^2 \mathbf{R} + 2\mathbf{K} \overline{\mathbf{y}'^o \mathbf{x}'^b} (1 - \mathbf{K}) \\ &= \zeta_1 + \zeta_2 + \zeta_3.\end{aligned}\tag{8}$$

The distribution of this quantity is shown by the dotted curve in Fig. 1a. Note that the EnKF \mathbf{P}^a distribution has the same mean (~ 0.44) as the EnKF-noPO distribution, since it suffers from the same problem associated with the nonlinearity in the covariance update. However, the distribution of \mathbf{P}^a for the EnKF is broader than that for the EnKF-noPO, indicating that the sampling error is larger and the analysis-error covariance estimate is less accurate, on average. Indeed, the mean absolute error of the estimated \mathbf{P}^a is ~ 0.24 for the EnKF, as compared to ~ 0.14 for the EnKF-noPO. Fig. 2 shows the mean absolute error of the estimated \mathbf{P}^a for the EnKF and EnKF-noPO as a function of ensemble size. The horizontal line on this figure highlights the fact that a 13-member EnKF ensemble is necessary to obtain an analysis-error covariance estimate that is as accurate as that obtained from a five-member EnKF-noPO ensemble.

From Fig. 1a, it is also clear that very small values of \mathbf{P}^a are significantly more probable after an application of the EnKF with perturbed observations than after an application of a hypothetical EnKF-noPO. As a consequence, there is a higher probability that the estimated \mathbf{P}^a will be less than the exact value when observations are perturbed (62% vs. 59% for the EnKF-noPO). To understand this, let us consider the contribution of each term in Eq. 8. Since these terms all contain \mathbf{x}'^b , they are not independent and one cannot infer the probability distribution of \mathbf{P}^a ($P(\mathbf{P}^a)$) from the unconditional distributions of ζ_1, ζ_2 , and ζ_3 . Fig. 1b shows the dependence of $P(\zeta_3)$ on $\zeta_1 + \zeta_2$. When $\zeta_1 + \zeta_2$ is large, so is the variance of ζ_3 . The solid curve in Fig. 1c shows $P(\zeta_1 + \zeta_2)$ after the assimilation of the perturbed observations. Over-plotted on this are curves of $P(\zeta_3)$ for two different values of $\zeta_1 + \zeta_2$. $P(\zeta_3)$ is broader when $\zeta_1 + \zeta_2$ is larger, and the probability of adding noise with a certain variance is proportional to the probability of a given $\zeta_1 + \zeta_2$.

Consider the effect of the noise term ζ_3 on the probability of \mathbf{P}^a having the value indicated at the point labeled “a.” The noise term ζ_3 in \mathbf{P}^a when $\zeta_1 + \zeta_2$ has the value indicated at point “b” spreads out probability, increasing $P(\mathbf{P}^a)$ at “a.” However, when $\zeta_1 + \zeta_2$ has the value given at “a,” the noise term does not increase $P(\mathbf{P}^a)$ at “b” in equal measure. Hence, the addition of noise from term ζ_3 can change the shape of the distribution of \mathbf{P}^a , further skewing the distribution of \mathbf{P}^a if the distribution of $\zeta_1 + \zeta_2$ is already skewed.

Houtekamer and Mitchell (1998) introduced the concept of a double EnKF, in which two ensemble data assimilation cycles are run in parallel with the background-error covariances from one ensemble being used to update the other. The double EnKF was designed to combat the biases introduced by the nonlinear dependence of the analysis-error covariance

on the background-error covariance. We have tested the double EnKF in this simple scalar model and have found that, as predicted by van Leeuwen (1999), it actually reverses the sign of this bias, resulting in an ensemble whose mean analysis-error variance is an overestimate of the “exact” value (which would be obtained with an infinite ensemble) but is a very accurate estimate of the actual ensemble-mean analysis error. However, the effects of the extra sampling error associated with perturbed observations are not mitigated by the double EnKF, and the mean absolute error in the estimation of \mathbf{P}^a for the double EnKF with $2n$ members is very similar to the single EnKF with n members (Fig. 2).

In summary, this simple example shows how perturbed observations can increase the sampling error of an ensemble data assimilation system. Pham (2001) introduced a method for perturbing observations which guarantees that the background-observation error covariance terms in Eq. (7) is zero. Under these circumstances, the deleterious effects of the perturbed observations just discussed should vanish. However, the computational cost of constraining the perturbed observations in this way would be significant in a model with many degrees of freedom. Therefore, an ensemble data assimilation system that does not require perturbed observations to maintain sufficient ensemble variance is desirable, and should be both more accurate and less susceptible to filter divergence than the traditional EnKF with unconstrained perturbed observations.

3. An ensemble square-root filter

We now define a revised EnKF that eliminates the necessity to perturb the observations. In this case, Eq (5) simplifies to $\mathbf{x}'^a = \mathbf{x}'^b - \tilde{\mathbf{K}}\mathbf{H}\mathbf{x}'^b = (\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})\mathbf{x}'^b$. We seek a definition for $\tilde{\mathbf{K}}$

that will result in an ensemble whose analysis-error covariance satisfies Eq. (3). Substituting $\tilde{\mathbf{K}}$ into Eq. (6), and requiring that the resulting expression be equal to the “correct” \mathbf{P}^a (given by Eq. (3)), yields an equation for $\tilde{\mathbf{K}}$

$$(\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})\mathbf{P}^b(\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})^T = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b \quad (9)$$

which has a solution

$$\tilde{\mathbf{K}} = \mathbf{P}^b\mathbf{H}^T \left[(\sqrt{\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}})^{-1} \right]^T (\sqrt{(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})} + \sqrt{\mathbf{R}})^{-1}, \quad (10)$$

(Andrews 1968). Since the calculation of $\tilde{\mathbf{K}}$ involves the square root of $p \times p$ error-covariance matrices (where p is the total number of observations), this is essentially a Monte-Carlo implementation of a square-root filter (Maybeck 1979). For this reason, we call this algorithm the ensemble square-root filter, or EnSRF. The matrix square roots in Eq. (10) are not unique, they can be computed in different ways, such as Cholesky factorization or singular value decomposition. Square-root filters were first developed for use in small computers on-board aircraft, where word-length restrictions required improved numerical precision and stability. These requirements outweighed the computational overhead required to compute the square-root matrices, which can be significant.

For an individual observation, $\mathbf{H}\mathbf{P}^b\mathbf{H}^T$ and \mathbf{R} are scalars, and Eq. (9) may be written

$$\frac{\mathbf{H}\mathbf{P}^b\mathbf{H}^T}{\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}} \tilde{\mathbf{K}}\tilde{\mathbf{K}}^T - \mathbf{K}\tilde{\mathbf{K}}^T - \tilde{\mathbf{K}}\mathbf{K}^T + \mathbf{K}\mathbf{K}^T = \mathbf{0}. \quad (11)$$

If $\tilde{\mathbf{K}} = \alpha \mathbf{K}$, where α is a constant, then $\mathbf{K}\mathbf{K}^T$ may be factored out of the above equation, resulting in a scalar quadratic for α ,

$$\frac{\mathbf{H}\mathbf{P}^b\mathbf{H}^T}{\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}}\alpha^2 - 2\alpha + 1 = 0. \quad (12)$$

Since we want the deviations from the ensemble mean to be reduced in magnitude while maintaining the same sign, we choose the solution to Eq. (12) that is between 0 and 1. This solution is

$$\alpha = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}}}\right)^{-1}. \quad (13)$$

This was first derived by Potter (1964). Here, $\mathbf{H}\mathbf{P}^b\mathbf{H}^T$ and \mathbf{R} are scalars representing the background and observational error variance at the observation location. Using $\tilde{\mathbf{K}}$ to update deviations from the mean, the analysis error covariance is guaranteed to be exactly equal to Eq. (3). In the EnKF, the excess variance reduction caused by using \mathbf{K} to update deviations from the mean is compensated for by the introduction of noise to the observations. In the EnSRF, the mean and departures from the mean are updated independently according to Eqns. (2), (4), (5) and (13), with $\mathbf{y}'^o = \mathbf{0}$. If \mathbf{R} is diagonal, observations may be processed one at a time, and the EnSRF requires no more computation than the traditional EnKF with perturbed observations. In fact, as noted by Maybeck (1979; p. 375) the most efficient way of updating the deviations from the mean in the square-root filter is to process the measurements one at a time using the Potter algorithm described above, due to the extra overhead of computing matrix square roots when observations are processed in batches. If

observation errors are correlated (\mathbf{R} is not diagonal), the variables may be transformed into the space in which \mathbf{R} is diagonal, or observations may be with correlated errors may be processed in batches using Eq. (10).

The Kalman filter equations (Eqns. 1, 2, and 3) are derived assuming that the forward operator \mathbf{H} , which interpolates the first-guess to the observations locations, is linear. In an operational setting, \mathbf{H} may be a nonlinear operator, such as a radiative transfer model that converts temperature, humidity and ozone concentrations into radiances. Houtekamer and Mitchell (2001) noted that, since in the EnKF \mathbf{H} is applied to the background fields individually (instead of to the covariance matrix \mathbf{P}^b directly), it is possible to use a nonlinear \mathbf{H} . However, using a nonlinear \mathbf{H} (as opposed to a linearized \mathbf{H}) in the EnKF may not necessarily improve the analysis, since it violates the assumptions used to derive the analysis equations. If, as suggested by Houtekamer and Mitchell (2001), \mathbf{H} is applied to the background fields individually, before computation of $\mathbf{P}^b\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}^b\mathbf{H}^T$, then the derivation of Eq. (13) applies even when \mathbf{H} is nonlinear. However, all of the results presented in this paper are for linear \mathbf{H} ; we have not tested the behavior of the EnSRF for nonlinear \mathbf{H} .

The ensemble adjustment filter (EnAF) of Anderson (2001), the error-subspace statistical estimation method (ESSE) of Lermusiaux and Robinson (1999), and the ensemble transform Kalman filter (ETKF) of Bishop et al. (2001), all use the traditional Kalman filter update equation (Eq. (1)) to update the ensemble mean, and constrain the updated ensemble covariance to satisfy Eq. (3). Therefore, they are all functionally equivalent, differing only in algorithmic details. In the EnAF, the linear operator $\mathbf{A} = \mathbf{I} - \tilde{\mathbf{K}}\mathbf{H}$ is applied to the prior, or background, ensemble to get an updated ensemble whose sample covariance is identical

to that computed by Eq. (3). In the ESSE, Eq. (3) is used explicitly to update the eigenvectors and eigenvalues of \mathbf{P}^b . A random sample of the updated eigenspectrum is then used as initial conditions for the next ensemble forecast. In the ETKF, the matrix square roots are computed in the subspace spanned by the ensemble. Here we have shown that all of these methods have a common ancestor, the square-root filter developed in the early 1960's for aeronautical guidance systems with low precision. By applying the algorithm of Potter (1964), we have presented a computationally simple way of constraining the updated ensemble covariance which is no more expensive than the traditional EnKF when observations are processed serially, one at a time. Sequential processing of observations also makes it much easier to implement covariance localization, which improves the analysis while preventing filter divergence in small ensembles (Houtekamer and Mitchell 2001; Hamill et al. 2001).

Burgers et al. (1998) showed that the covariance estimates produced by EnKF with perturbed observations asymptote to traditional extended Kalman filter result as the number of ensemble members tends to infinity. For ensemble data assimilation schemes based on the square-root filter without perturbed observations, the extended Kalman filter result is obtained when the number of ensemble members equals the number of state variables (C. Bishop, personal communication). This implies that in such cases the the ensemble and the true state cannot be considered to be drawn from the same probability distribution. However, in most real-world applications this is not an issue since the number of ensemble members is likely to be far less than the number of state variables.

4. Results with the 40-variable Lorenz model

The simple example presented in section 2b demonstrated that perturbing the observations in the EnKF can have a potentially serious impact on filter performance. In this section we attempt to quantify this impact by comparing results with the EnKF and EnSRF in the 40-variable dynamical system of Lorenz and Emanuel (1998), and then in a idealized general circulation model with roughly 2000 state variables in section 5.

The model of Lorenz and Emanuel (1998) is governed by the equation

$$\frac{dX_i}{dt} = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F, \quad (14)$$

where $i = 1, \dots, m$ with cyclic boundary conditions. Here we use $m = 40$, $F = 8$ and a fourth-order Runge-Kutta time integration scheme with a time step of 0.05 units. For this parameter setting, the leading Lyapunov exponent implies an error-doubling time of about 8 time steps, and the fractal dimension of the attractor is about 27 (Lorenz and Emanuel 1998). For our assimilation experiments, each state variable is observed directly, and observations have uncorrelated errors with unit variance. A 10-member ensemble is used, and observations are assimilated every time step for 50000 time steps (after a spin-up period of 1000 time steps). Observations are processed serially (one after another) for the EnSRF, and simultaneously (by inverting a 40×40 matrix) for the EnKF. The RMS error of the ensemble mean (E_1) is defined as

$$E_1 = \sqrt{\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{n} \sum_{j=1}^n X_i^j - X_i^{true} \right)^2}, \quad (15)$$

where $n = 10$ is the number of ensemble members, $m = 40$ is the number of state variables,

X_i^j is the j th ensemble member for the i th variable, and X^{true} is the “true” state from which the observations were sampled. E_2 is the average RMS error of each ensemble member,

$$E_2 = \frac{1}{n} \sum_{j=1}^n \sqrt{\frac{1}{m} \sum_{i=1}^m (X_i^j - X_i^{true})^2}. \quad (16)$$

The RMS ratio $R = E_1/E_2$ is a measure of how similar the truth is to a randomly selected member of the ensemble (Anderson 2001). If the truth is statistically indistinguishable from any ensemble member then the expected value of $R = \sqrt{(n+1)/2n}$, or approximately 0.74 for a 10-member ensemble (Murphy 1988). When $R < 0.74$ ($R > 0.74$) the ensemble variance overestimates (underestimates) the error of the ensemble mean (E_1).

As discussed in section 2b, sampling error can cause filter divergence in any ensemble data assimilation system, so some extra processing of the ensemble covariances is almost always necessary. The two techniques used here are distance-dependent covariance filtering (Houtekamer and Mitchell 2001; Hamill et al. 2001) and covariance inflation (Anderson and Anderson 1999).

Covariance filtering counters the tendency for ensemble variance to be excessively reduced by spurious long-range correlations between analysis and observations points by applying a filter that forces the ensemble covariances to go to zero at some distance L from the observation being assimilated. In the EnSRF some care must be taken when applying covariance localization, since the derivation of Eq. (13) assumes that there has been no filtering of the background-error covariances used to compute the Kalman gain. In fact, if covariance localization is applied in the computation of \mathbf{K} and Eq. (13) is used to update the ensemble deviations, the analysis-error covariances will only satisfy Eq. (3) if the covariance

localization function is a Heaviside step function, i.e. the background-error covariances are set to zero beyond a certain distance L from the observation and are unmodified otherwise. However, we have found that using smoother covariance localization functions, such as the function given by Eq. 4.10 in Gaspari and Cohn (1999), while not strictly constraining the ensemble to satisfy Eq. (3), actually produce a more accurate analysis than the Heaviside function. Hence, all of the results presented here use the Gaspari-Cohn function for covariance localization.

The nonlinear dependence of the gain matrix on the background-error covariance will result in a biased estimate of analysis error, and the analysis system will weight the first-guess forecasts too heavily on average. Underestimation of the background-error covariances has a more severe impact on analysis error than overestimation (Daley 1991, section 4.9). To compensate for this, covariance inflation simply inflates the deviations from the ensemble mean first-guess by a small constant factor r for each member of the ensemble, before the computation of the background-error covariances and before any observations are assimilated. The double EnKF introduced by Houtekamer and Mitchell (1998) combats this problem by using parallel ensemble data assimilation cycles in which the background-error covariances estimated by one ensemble are used to calculate the gain for the other. In the Appendix, we develop a double EnSRF and present results as a function of covariance filter length scale for the 40-variable model. The results are qualitatively similar to those shown here for the “single” filters with covariance inflation. However, as noted in the Appendix, the double EnSRF requires significantly more computation than the single EnSRF, since the simplifications to Eq. (9) obtained for serial processing of observations are no longer applicable.

Fig. 3 shows E_1 averaged over 50000 assimilation cycles for the EnKF and EnSRF, as a function of the covariance inflation factor and the length scale of the covariance localization filter. The shaded areas on these plots indicate regions in parameter space where the filter has diverged, i.e. has drifted into a regime where it effectively ignores observations. For both the EnKF and EnSRF filter divergence occurs for a given covariance filter length scale L when r is less than a critical value. However, the EnSRF appears to be less susceptible to filter divergence, since the critical value of the covariance inflation factor is always less for a given L . Overall, for almost any parameter value, the error in the EnSRF is less than the EnKF. The minimum error in the EnSRF is 0.2, which occurs at $L = 24$ for $r = 1.03$. For the EnKF, the minimum error is 0.26, which occurs at $L=15$ for $r = 1.07$. These results are consistent with those presented for the simple scalar example in section 2b. The observational noise terms in Eq. 7 increases the sampling error in the estimation of the error covariances in the EnKF, so that a larger value of the covariance inflation factor r is needed to combat filter divergence. The reduced sampling error in the EnSRF allows the algorithm to extract more useful covariance information from the ensemble for a given ensemble size. This is consistent with the fact that larger values of L benefit the EnSRF, but are detrimental to the EnKF.

The results of section 2b suggest that, because of reduced sampling error, the EnSRF should produce a more accurate estimate of analysis-error variance than the EnKF for a given ensemble size. Therefore, if both the EnKF and the EnSRF start with the same \mathbf{P}^b , the EnSRF should produce a more accurate \mathbf{P}^a when the ensemble is updated with observations. After propagating the resulting ensembles forward to the next analysis time with a model, the EnSRF ensemble should therefore yield a more accurate estimate of \mathbf{P}^b , thus producing

an ensemble mean with lower error after the next set of observations are assimilated, even though the expected analysis-error variance is the same as obtained with the EnKF. The ratio of ensemble mean error to analysis-error variance should therefore be larger in the EnKF than the EnSRF. Fig. 4 shows the RMS ratio for the 40-variable model experiments. For a 10-member ensemble, the expected value for an ensemble which faithfully represents the true underlying probability distribution is 0.74. If the RMS ratio is smaller (larger) than this value, the ensemble overestimates (underestimates) the uncertainty in the analysis. For nearly all parameter settings the EnSRF has a lower RMS ratio than the EnKF, indicating that, as expected, the variance in the EnKF ensemble is indeed smaller relative to ensemble mean analysis error.

5. Results with an idealized two-level GCM

The results shown thus far have been for very simple models, where the number of ensemble members, the number of observations and the number of model variables are all of the same order. In this section we will present results with an idealized primitive equation GCM with roughly 2000 degrees of freedom, in order to demonstrate that these results also apply to more realistic higher-dimensional systems.

The model is virtually identical to the two-level (250 and 750 hPa) model described by Lee and Held (1993), run at T31 horizontal resolution with hemispheric symmetry imposed. The prognostic variables of the model are baroclinic and barotropic vorticity, baroclinic divergence and barotropic potential temperature, yielding 2047 state variables in all. Barotropic divergence is assumed to be zero, and the baroclinic potential temperature is set to a con-

stant value of 10 K . The lower level winds are subjected to a mechanical damping with a time scale of 4 days, and the baroclinic potential temperature is relaxed back to a radiative equilibrium state with a pole-to-equator temperature difference of 80 K with a time scale of 20 days. The functional form of the radiative equilibrium profile is as given by Eq. 3 in Lee and Held (1993). A quad-harmonic (∇^8) diffusion is applied to all prognostic variables, and the coefficient is chosen so that the smallest resolvable scale is damped with a e-folding time scale of three hours. An explicit fourth-order Runge-Kutta time integration scheme is used, with 14 model time steps per day. Based on the leading Lyapunov exponent, the error doubling time of this model is about 2.4 days.

Starting with a random perturbation superimposed upon a resting state, the model is integrated for 400 days. The first 200 days are discarded. Observations are generated by sampling the model output at selected locations every 12 hours, with random noise added to simulate observation error. There are 46 observation locations, defined by a geodesic grid such that points are roughly equally spaced on the surface of the hemisphere at an interval of approximately 2300 km (Fig. 5). At each observation location, there are observations of winds at 250 and 750 hPa and temperature at 500 hPa. The error standard deviations for wind and temperature are set to 1 ms^{-1} and 1 K , respectively. 20-member ensembles are used, and the initial ensemble consists of randomly selected model states from the reference integration. Assimilation experiments are run for 180 days, and statistics are computed over the last 120 days. Experiments were conducted using the EnSRF and the EnKF with perturbed observations for a range of covariance filter length scales and covariance inflation factors. Observations were processed serially in both the EnSRF and the EnKF. EnKF ex-

periments with simultaneous observation processing for selected parameter settings produced nearly identical results, so serial processing was used because it runs faster and requires less memory.

Figs. 6 and 7 summarize the results for 500 hPa potential temperature RMS error and RMS ratio, as a function of covariance filter length scale and covariance inflation factor. The gray shaded areas indicate filter divergence. The results are qualitatively similar to those obtained with the 40-variable Lorenz model. The minimum RMS error for the EnSRF is 0.096 K , which occurs for a covariance filter length scale of 9500 km and a covariance inflation factor of 1.05. For the EnKF, the minimum error is 0.125 K , which occurs at $L = 7500$ km and $r = 1.10$. RMS ratios are generally larger for the EnKF indicating that relative to the EnSRF, ensemble variance is smaller relative to ensemble mean error. The EnSRF also is less susceptible to filter divergence than the EnKF, also in agreement with the Lorenz 40-variable results.

6. Conclusions

If the same gain and the same observations are used to update each ensemble member in an ensemble data assimilation system, analysis-error covariances will be systematically underestimated, and filter divergence may occur. To overcome this, an ensemble of perturbed observations can be assimilated, whose statistics reflect the known observation errors. In the limit of infinite ensemble size, in a system where all sources of error (both observation and model) are correctly sampled, this approach yields the correct analysis-error covariances. However, when ensemble sizes are finite, the noise added to the observations produces spu-

rious observation-background error covariances associated with sampling error in the estimation of the observation-error covariances. Using a very simple example, we demonstrate how sampling error in both the background and observational error covariances can affect the accuracy of analysis-error covariance estimates in ensemble data assimilation systems. Because of the nonlinear dependence of the analysis-error covariance on the background-error covariance, sampling error in the estimation of background-error covariances will cause analysis-error covariances to be underestimated on average. Thus, techniques to boost ensemble variance are almost always necessary in ensemble data assimilation to prevent filter divergence. The addition of noise to the observations in an ensemble data assimilation system has two primary effects; (1) reducing the accuracy of the analysis-error covariance estimate by increasing the sampling error, (2) increasing the probability that the analysis-error covariance will be underestimated by the ensemble. The latter is a consequence of the fact that the perturbed observations act as a source of multiplicative noise in the analysis system, since the amplitude of the noise in the analysis-error covariances is proportional to the background error estimate. Underestimation of error covariances is undesirable, and has a larger impact on analysis error than a commensurate overestimation. Given these effects, we expect that ensemble data assimilation methods which use perturbed observations should have a higher error for a given ensemble size than a comparable system without perturbed observations. Designing an ensemble data assimilation system which does not require perturbed observations is therefore a desirable goal.

Various techniques for ensemble data assimilation which do not require perturbing the observations have recently been proposed. We have designed and tested an algorithm which

we call the “ensemble square-root filter”, or EnSRF. The algorithm avoids the systematic underestimation of the posterior covariance which led to the use of perturbed observations in EnKF by using a different Kalman gain to update the ensemble mean and deviations from the ensemble mean. The traditional Kalman gain is used to update the ensemble mean, and a “reduced” Kalman gain is used to update deviations from the ensemble mean. The reduced Kalman gain is simply the traditional Kalman gain times a factor between 0 and 1 which depends only on the observation and background error variance at the observation location. This scalar factor is determined by the requirement that the analysis-error covariance match what would be predicted by the traditional Kalman filter covariance update equation (Eq. 3) given the current background-error covariance estimate. The EnSRF algorithm proposed here involves processing the observations serially, one at a time. This implementation is attractive because, in addition to being algorithmically simple, it avoids the need to compute matrix square roots and thus requires no more computation than the EnKF with perturbed observations and serial observation processing.

The benefits of ensemble data assimilation without perturbed observations are demonstrated by comparing the EnKF and the EnSRF in a hierarchy of models, ranging from a simple scalar model to a idealized primitive equation GCM with $O(10^3)$ state variables. In all cases, the EnSRF produces an analysis ensemble whose ensemble mean error is lower than the EnKF for the same ensemble size.

The EnSRF as formulated here requires observations to be processed one at a time, which may pose quite a challenge in an operational setting where observations can number in the millions. It will be crucial to develop algorithms that allow observations at locations where

background errors are uncorrelated to be processed in parallel. The treatment of model error, which we have not considered in this study, will likely be a crucial element in any future operational system. These will continue to be active areas of research as ensemble data assimilation methods are implemented in more complex and realistic systems.

Acknowledgments

Fruitful discussions with Jeffrey Anderson, Thomas Bengtson, Gilbert Compo Doug Nychka, Chris Snyder and Michael Tippett are gratefully acknowledged, as are the detailed and thoughtful reviews by Craig Bishop and Peter Houtekamer.

References

- Anderson, J. L., 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.
- Andrews, A., 1968: A square root formulation of the Kalman covariance equations. *AIAA J.*, **6**, 1165–1168.
- Bishop, C. H., B. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.*, **129**, 420–436.

- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Daley, R., 1991: *Atmospheric Data Assimilation*. Cambridge University Press, 457 pp.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99 (C5)**, 10143–10162.
- Evensen, G. and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85–96.
- Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.
- Gelb, A., J. F. Kasper, R. A. Nash, C. F. Price, and A. A. Sutherland, 1974: *Applied Optimal Estimation*. M. I. T. Press, 374 pp.
- Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, in press.
- Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- 1999: Reply to comment on "Data assimilation using an ensemble Kalman filter". *Mon. Wea. Rev.*, **127**, 1374–1385.

- 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.
- Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: operational, sequential, and variational. *J. Met. Soc. Japan*, **75 (1B)**, 181–189.
- Kalman, R. and R. Bucy, 1961: New results in linear prediction and filtering theory. *Trans. AMSE J. Basic Eng.*, **83D**, 95–108.
- Lee, S. and I. M. Held, 1993: Baroclinic wave packets in models and observations. *J. Atmos. Sci.*, **50**, 1413–1428.
- Lermusiaux, P. F. J. and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.*, **127**, 1385–1407.
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.
- Lorenz, E. N. and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414.
- Maybeck, P. S., 1979: *Stochastic Models, Estimation and Control*, Academic Press, volume 1, chapter 7. 368–409.
- Murphy, J. M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.*, **116**, 89–125.
- Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207.

- Potter, J., 1964: W matrix augmentation. M.I.T. Instrumentation Laboratory Memo SGA 5-64, Massachusetts Institute of Technology, Cambridge, MA.
- van Leeuwen, P. J., 1999: Comment on "Data assimilation using an ensemble Kalman filter technique". *Mon. Wea. Rev.*, **127**, 1374–1377.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.

Appendix: The double EnSRF.

As discussed in section 2, in any ensemble data assimilation system the nonlinear dependence of the gain matrix on the background-error covariance will result in a biased estimate of analysis-error covariances, and the analysis system will weight the first-guess forecasts too heavily on average. Houtekamer and Mitchell (1998) referred to this problem as the “inbreeding effect”. Left unchecked, this problem may eventually lead to a condition called “filter divergence”, in which the ensemble spread decreases dramatically while the ensemble mean error increases and the analysis drifts farther and farther from the true state. To prevent this, covariance inflation (Anderson and Anderson 1999) simply inflates the deviations from the ensemble mean first-guess by a small constant factor for each member of the ensemble, before the computation of the background-error covariances and before any observations are assimilated. The “forgetting factor”, used by Pham (2001), plays essentially the same role. Another way to deal with this problem, the so-called “double EnKF” was introduced by Houtekamer and Mitchell (1998). The double EnKF uses parallel ensemble data assimila-

tion cycles in which the background-error covariances estimated by one ensemble are used to calculate the Kalman gain for the other. van Leeuwen (1999) presents an analysis of the double EnKF which shows how it compensates for the biases associated with nonlinearity in the Kalman gain. Here we present a comparison between a double EnSRF and the double EnKF to complement the results for the single filters with covariance inflation presented in section 4.

In the double EnKF, the expected value of the analysis-error covariance for an infinite ensemble is

$$\mathbf{P}_i^a = \mathbf{P}_i^b + \mathbf{K}_{3-i}(\mathbf{H}\mathbf{P}_i^b\mathbf{H}^T + \mathbf{R})\mathbf{K}_{3-i}^T - \mathbf{P}_i^b\mathbf{H}^T\mathbf{K}_{3-i}^T - \mathbf{K}_{3-i}\mathbf{H}\mathbf{P}_i^b, \quad (\text{A1})$$

(van Leeuwen 1999), where the subscripts $i, 3-i$ ($i = 1, 2$) denote which ensemble the quantity was calculated from. Therefore, the counterpart of Eq. (9) for the double EnSRF is

$$(\mathbf{I} - \widetilde{\mathbf{K}}_i\mathbf{H})\mathbf{P}_i^b(\mathbf{I} - \widetilde{\mathbf{K}}_i\mathbf{H})^T = \mathbf{K}_{3-i}(\mathbf{H}\mathbf{P}_i^b\mathbf{H}^T + \mathbf{R})\mathbf{K}_{3-i}^T - \mathbf{P}_i^b\mathbf{H}^T\mathbf{K}_{3-i}^T - \mathbf{K}_{3-i}\mathbf{H}\mathbf{P}_i^b. \quad (\text{A2})$$

As before, if observations are processed one at a time, $\mathbf{H}\mathbf{P}_i^b\mathbf{H}^T$ and \mathbf{R} are scalars. However, substituting $\widetilde{\mathbf{K}}_i = \boldsymbol{\alpha}\mathbf{K}_i$ does not result in a scalar quadratic equation for $\boldsymbol{\alpha}$, since $\mathbf{K}_i\mathbf{K}_i^T$ does not factor out of the equation. Therefore, even when observations are processed serially, one at a time, the computation of $\widetilde{\mathbf{K}}_i$ involves the computation of matrix square-roots. Since this negates the advantage of serial processing, we elect to process all the observations at once. Defining $\mathbf{A}_i = \mathbf{I} - \widetilde{\mathbf{K}}_i\mathbf{H}$ results in the following equation for \mathbf{A}_i ,

$$\mathbf{A}_i = \sqrt{\mathbf{P}_i^b} \left[\sqrt{\mathbf{P}_i^a} \right]^{-1}, \quad (\text{A3})$$

where \mathbf{P}_i^a is given by Eq. (A1). This expression for \mathbf{A} is identical to the one derived in the Appendix of Anderson (2001). The matrix square-roots are computed using a lower-triangular Cholesky factorization. The double EnKF is solved as described by Houtekamer and Mitchell (1998).

Fig. A1a shows the ensemble mean RMS error for the double EnSRF and the double EnKF, as a function of covariance filter length scale L (the covariance inflation factor is not needed in the double filter) for 20 total ensemble members (10 in each ensemble). The experimental set-up is as described in section 4, except that two parallel data assimilation cycles are run. For all but the shortest L , the double EnSRF is more accurate than the double EnKF. The minimum error for the double EnSRF (EnKF) is 0.2 (0.25) for $L = 20$ ($L = 17$). This is quantitatively similar to the results obtained for the single filters with covariance inflation in section 4, suggesting that the conclusions and interpretations presented there are not sensitive to the method used to counter the tendency for filter-divergence associated with nonlinearity in the Kalman gain (i.e. the “inbreeding effect” noted by Houtekamer and Mitchell (1998)). Fig. A1b shows the ensemble mean RMS error for the double filters as a function of ensemble size. Values are plotted for ‘optimal’ covariance filter length scales, i.e. those values of L at which the error is a minimum. Clearly, the double EnSRF is more accurate for a given ensemble size than the double EnKF. The horizontal line in that figure is shown to illustrate the fact the a double EnSRF with a total of 14 members (7 in each ensemble) is as accurate as the double EnKF with 38 total members (19 in each ensemble). However, as formulated here, in addition to the extra computational overhead associated with running two 10-member ensembles instead of one, the double EnSRF requires

the computation of two $m \times m$ Cholesky lower-triangular matrix square-roots, as well as the inverse of a lower-triangular matrix, where m is the dimension of the model state vector. This makes the double EnSRF much more computationally expensive than either the double EnKF or the single EnSRF.

Figure Captions

Figure 1: (a) Probability distribution of background error variance (solid) and analysis error covariance (dotted for EnKF and dashed for EnKF-noPO) after assimilation of first observation. Solid vertical line indicates exact analysis error variance, dotted vertical line indicates mean of analysis error distribution after assimilation of first observation for EnKF and EnKF-noPO. See text for details. (b) Scatter-plot showing how distribution of ζ_3 is related to $\zeta_1 + \zeta_2$ in expression for \mathbf{P}_1^a (Eq. 8). (c) Illustration of how the multiplicative nature of the noise in ζ_3 skews the distribution of \mathbf{P}_1^a . Dashed line is the distribution of $\zeta_1 + \zeta_2$; shaded areas indicate the amount and the distribution of spurious background-perturbed observation noise added for two particular values of this quantity. Dotted line is probability distribution for sum of $\zeta_1 + \zeta_2 + \zeta_3$, i.e., \mathbf{P}_1^a .

Figure 2: Mean absolute error of estimated \mathbf{P}^a as a function of ensemble size n for the EnKF with perturbed observations (solid) and the EnKF-noPO (dotted). The horizontal line indicates the error of the EnKF-noPO estimate for a five-member ensemble. The dashed line shows the result for a 'double' EnKF with $2n$ total ensemble members (see text for details).

Figure 3: Ensemble mean error as a function of the distance at which the covariance filter goes to zero, and the covariance inflation factor, for the EnKF (A) and the EnSRF (B). Results are for a 10 member ensemble averaged over 5000 assimilation cycles using the model of Lorenz and Emanuel (1998), with observations of every state variable. Observations have unit error variance. Shaded regions indicate regions in parameter space where the filter diverges.

Figure 4: RMS ratio as a function of the distance at which the covariance filter goes to zero, and the covariance inflation factor, for the EnKF (A) and the EnSRF (B). Results are for a 10 member ensemble averaged over 5000 assimilation cycles using the model of Lorenz and Emanuel (1998), with observations of every state variable. Observations have unit error variance. Shaded regions indicate regions in parameter space where the filter diverges. The thicker contour is 0.74, which is the expected value for a 10 member ensemble.

Figure 5: 500 hPa potential temperature (contour interval 4 K) from the reference integration of the idealized two-level GCM with observation locations superimposed. The outermost latitude circle shown is $20^{\circ}N$.

Figure 6: As in Fig. 3, but for the idealized two-level primitive equation GCM described in the text.

Figure 7: As in Fig. 4, but for the idealized two-level primitive equation GCM described in the text. The thicker contour is 0.725, which is the expected value for a 20 member ensemble.

Figure A1: Double filter results for the 40-variable Lorenz model. (A) Ensemble mean error as a function of covariance filter length scale for the double EnSRF and double EnKF with 10 members in each ensemble. (B) Ensemble mean error as a function of ensemble size (as defined by the total number of members in both ensembles). Values in (B) are calculated with the covariance filter length scale L that yields the minimum error. The horizontal line in (B) denotes the error of the double EnSRF with 7 members in each ensemble (an ensemble size of 14). See the appendix for further details.

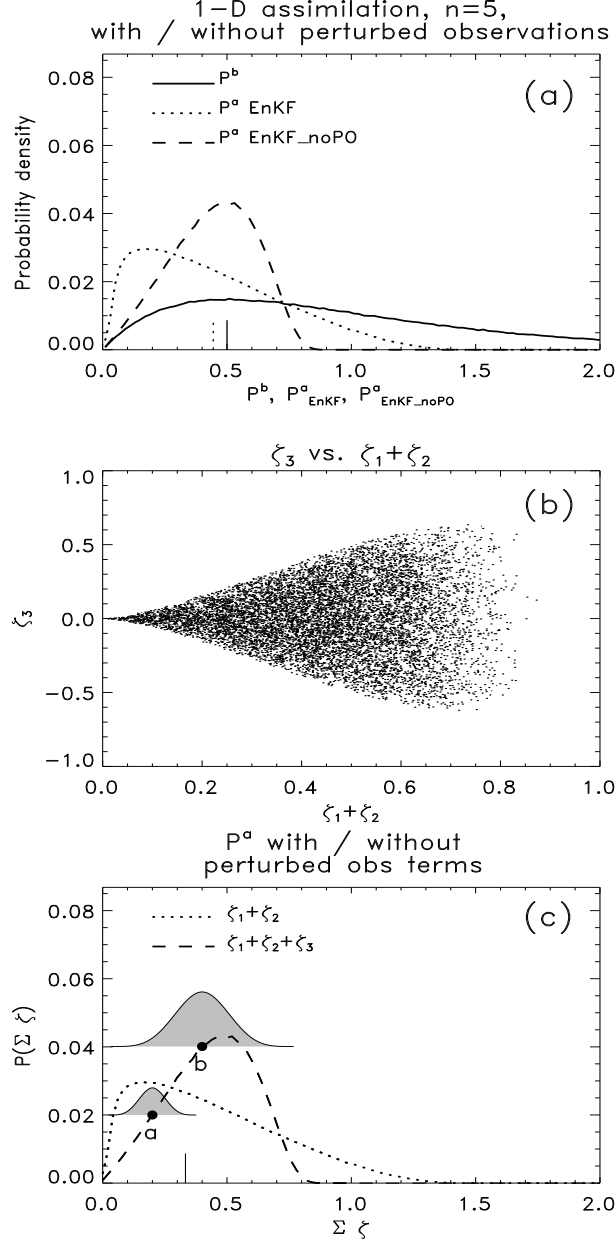


Figure 1: (a) Probability distribution of background error variance (solid) and analysis error covariance (dotted for EnKF and dashed for EnKF-noPO) after assimilation of first observation. Solid vertical line indicates exact analysis error variance, dotted vertical line indicates mean of analysis error distribution after assimilation of first observation for EnKF and EnKF-noPO. See text for details. (b) Scatter-plot showing how distribution of ζ_3 is related to $\zeta_1 + \zeta_2$ in expression for \mathbf{P}_1^a (Eq. 8). (c) Illustration of how the multiplicative nature of the noise in ζ_3 skews the distribution of \mathbf{P}_1^a . Dashed line is the distribution of $\zeta_1 + \zeta_2$; shaded areas indicate the amount and the distribution of spurious background-perturbed observation noise added for two particular values of this quantity. Dotted line is probability distribution for sum of $\zeta_1 + \zeta_2 + \zeta_3$, i.e., \mathbf{P}_1^a .

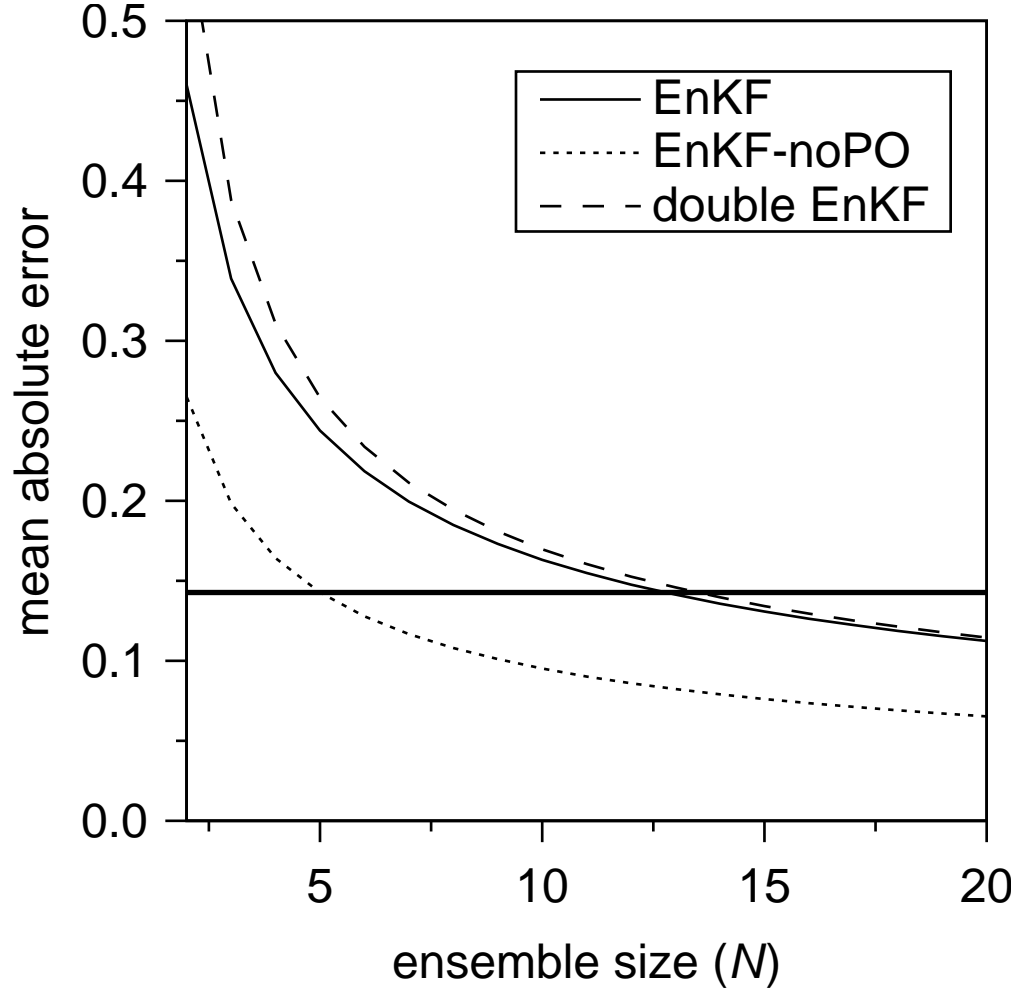


Figure 2: Mean absolute error of estimated \mathbf{P}^a as a function of ensemble size n for the EnKF with perturbed observations (solid) and the EnKF-noPO (dotted). The horizontal line indicates the error of the EnKF-noPO estimate for a five-member ensemble. The dashed line shows the result for a 'double' EnKF with $2n$ total ensemble members (see text for details).

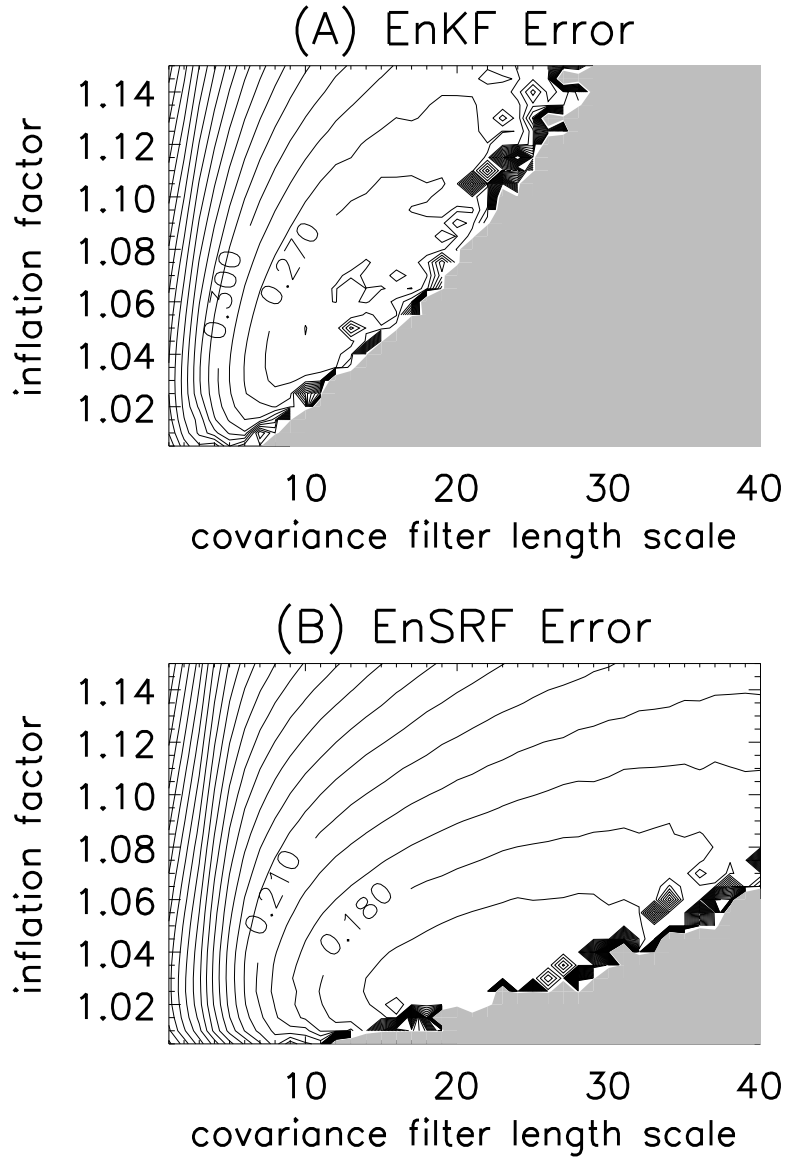


Figure 3: Ensemble mean error as a function of the distance at which the covariance filter goes to zero, and the covariance inflation factor, for the EnKF (A) and the EnSRF (B). Results are for a 10 member ensemble averaged over 5000 assimilation cycles using the model of Lorenz and Emanuel (1998), with observations of every state variable. Observations have unit error variance. Shaded regions indicate regions in parameter space where the filter diverges.

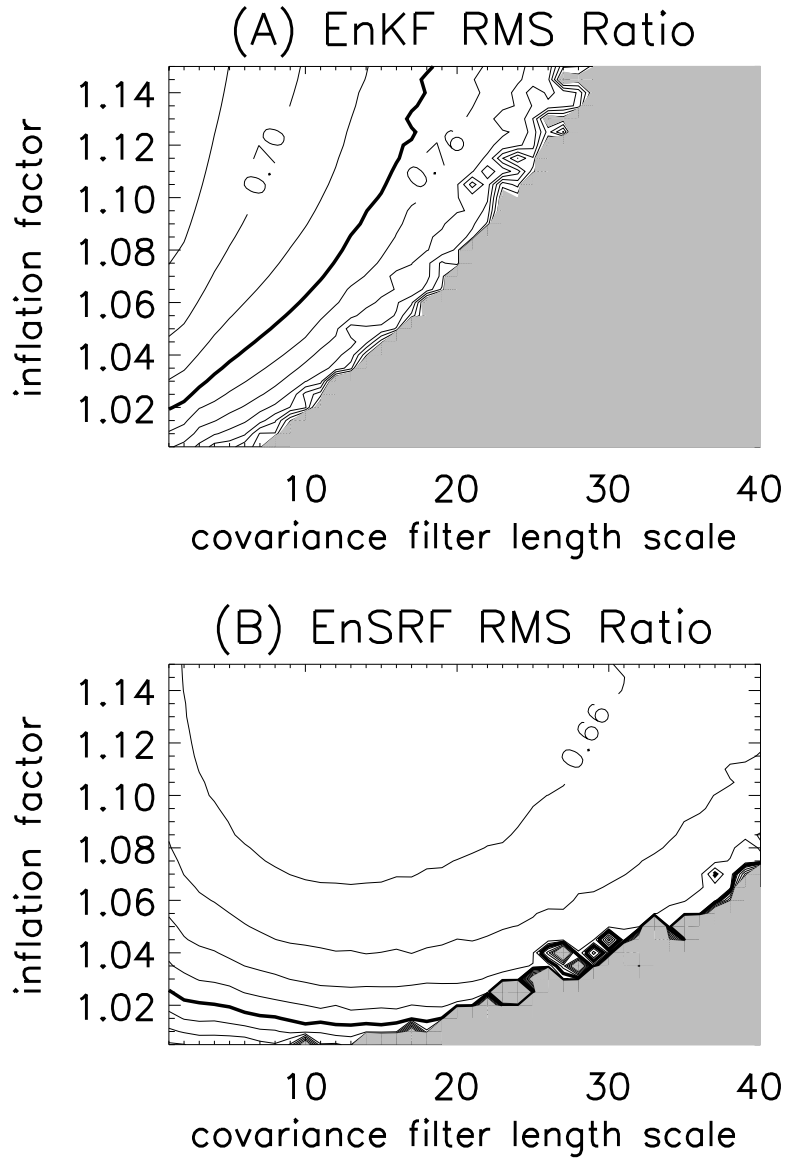


Figure 4: RMS ratio as a function of the distance at which the covariance filter goes to zero, and the covariance inflation factor, for the EnKF (A) and the EnSRF (B). Results are for a 10 member ensemble averaged over 5000 assimilation cycles using the model of Lorenz and Emanuel (1998), with observations of every state variable. Observations have unit error variance. Shaded regions indicate regions in parameter space where the filter diverges. The thicker contour is 0.74, which is the expected value for a 10 member ensemble.

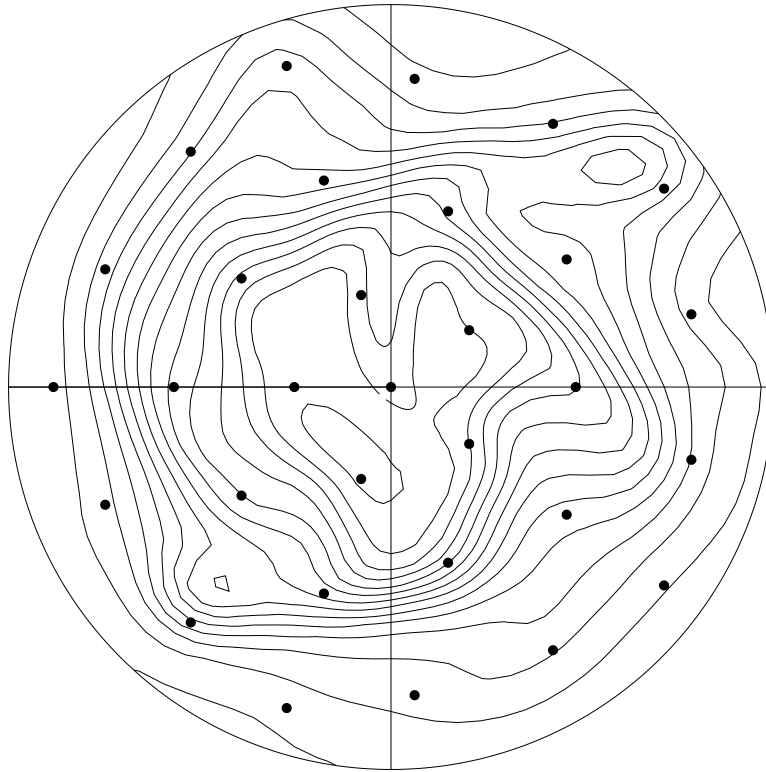


Figure 5: 500 hPa potential temperature (contour interval 4 K) from the reference integration of the idealized two-level GCM with observation locations superimposed. The outermost latitude circle shown is $20^{\circ}N$.

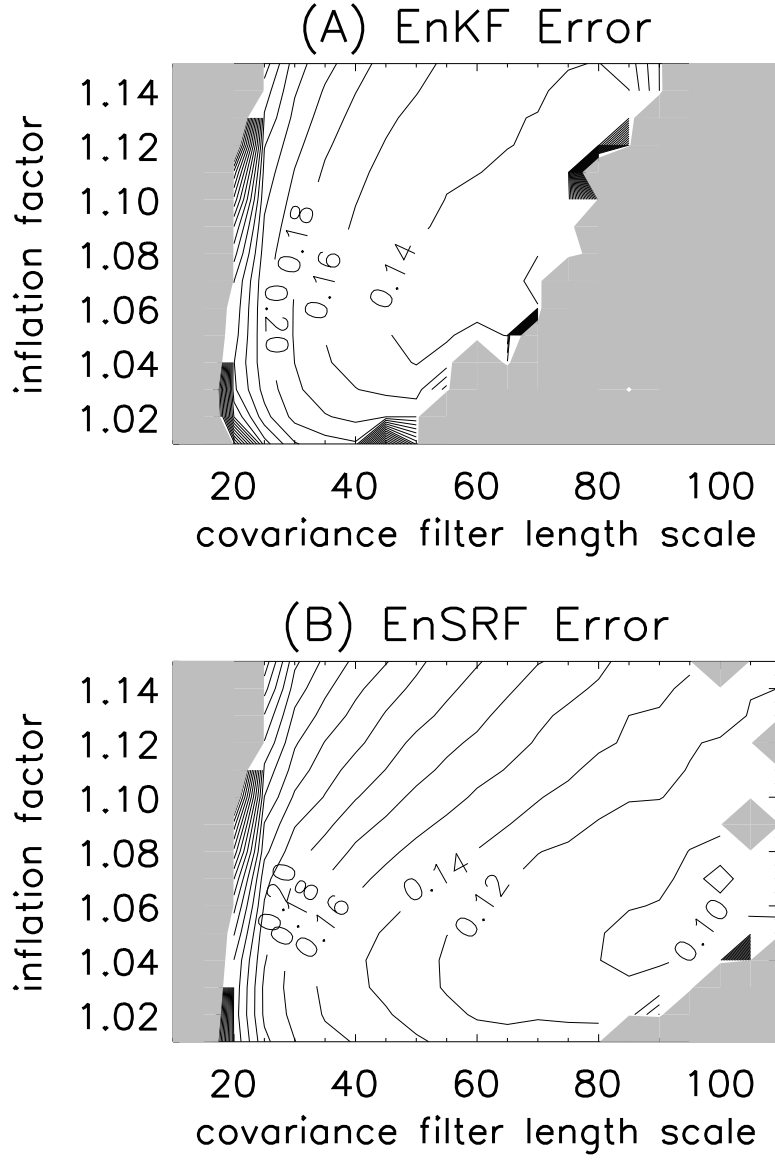


Figure 6: As in Fig. 3, but for the idealized two-level primitive equation GCM described in the text.

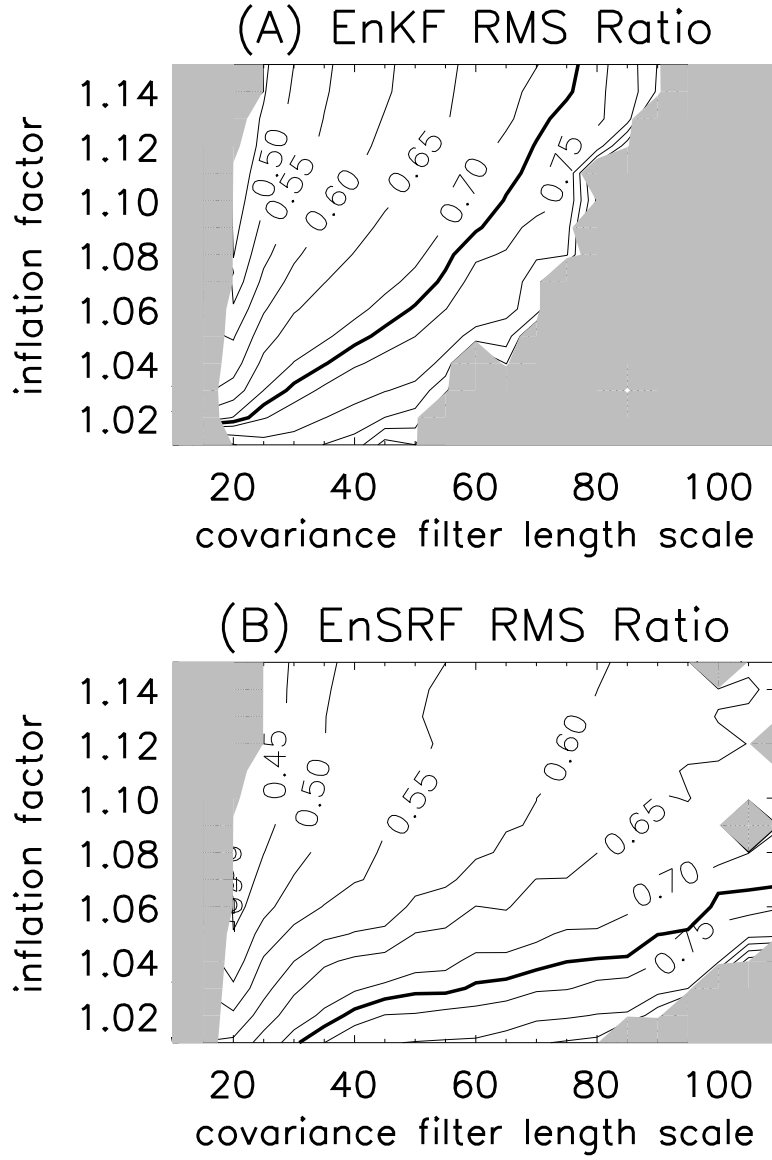


Figure 7: As in Fig. 4, but for the idealized two-level primitive equation GCM described in the text. The thicker contour is 0.725, which is the expected value for a 20 member ensemble.

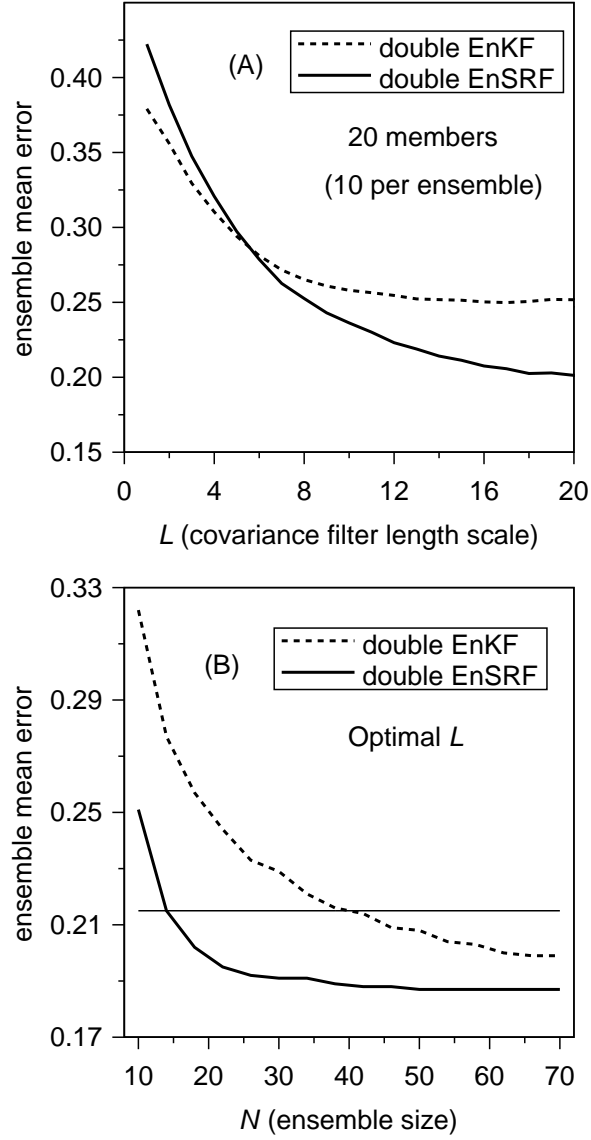


Figure A1: Double filter results for the 40-variable Lorenz model. (A) Ensemble mean error as a function of covariance filter length scale for the double EnSRF and double EnKF with 10 members in each ensemble. (B) Ensemble mean error as a function of ensemble size (as defined by the total number of members in both ensembles). Values in (B) are calculated with the covariance filter length scale L that yields the minimum error. The horizontal line in (B) denotes the error of the double EnSRF with 7 members in each ensemble (an ensemble size of 14). See the appendix for further details.